ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 10 (2012) 1140 – 1145
Procedia
Computer Science

The Second International Workshop on Sensor Networks for Intelligence Gathering and Monitoring

NOA: A Scalable Multi-parent Clustering Hierarchy for WSNs

Johnathan Vee Cree[a,b], Jose Delgado-Frias[a], Mike Hughes[b], Brion Burghard[b], Kurt Silvers[b]

[a] *School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA*
[b] *Pacific Northwest National Laboratory, Richland, WA*

Abstract

NOA is a multi-hop, multi-parent, N-tiered, hierarchical clustering algorithm that provides a scalable, robust and reliable solution for autonomous configuration of large-scale wireless sensor networks. The novel clustering hierarchy's inherent benefits can be utilized by in-network data processing techniques to provide a robust data processing solution capable of reducing the amount of data sent to data sinks. Utilizing a multi-parent framework, NOA reduces the cost of network configuration when compared to current hierarchical beaconing solutions by removing the r-hop flooding (where r is the radius of the cluster). NOA instead utilizes common children to distribute information about the hierarchy's topology to siblings. NOA^2, a two-parent clustering hierarchy solution, and NOA^3, the three-parent variant, saw up to an 83% and 72% reduction in communication overhead, respectively, when compared to configuring the network using a one-parent hierarchical beaconing solution, as well as 92% and 88% less overhead when compared to two- and three-parent variants of hierarchical beaconing.

Keywords:
wireless sensor networks, multi-parent clustering hierarchy, autonomous configuration, in-network data processing

1. Introduction

Wireless sensor networks (WSNs) for environmental monitoring have the primary objective of detecting abnormalities and extremes in an environment. The large-scale sensor networks needed for monitoring vast areas, in the absence of a significantly large number of data sinks, require multi-hop communication. As such these networks cannot rely on a centralized processing solution due to the communication cost associated with forwarding all of the data to a centralized location. In-network data processing can greatly reduce the amount of data sent to data sinks. When applied at multiple levels, in-network data processing can create spatiotemporal multi-resolution data sets for the analysis routines. As such, in-network analysis routines can focus on a range of resolutions from a brief overview of the entire network to a detailed description of the environment around a node. Further, many analysis routines require prior knowledge of the local environment with a similar spatiotemporal resolution. This leads to a natural tie between location and data aggregation/storage/analysis which when exploited can result in reducing the stored data even further.

Email addresses: `Johnathan.Cree@pnnl.gov` (Johnathan Vee Cree), `JDelgado@eecs.wsu.edu` (Jose Delgado-Frias)

1877-0509 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.procs.2012.06.163

1.1. Background

Monitoring WSNs generate copious amounts of redundant data that, due to communication costs, cannot all be forwarded to a data sink. As such, in-network data processing such as data aggregation [1], storage and query [2], anomaly detection [3], object tracking [4], etc. can greatly reduce the power cost of data communications by sending only the pertinent information to data sinks. NOA's novel multi-parent hierarchical structure provides in-network data processing applications with unique inherent benefits including but not limited to: multi-path communications, data redundancy, overlapping clustering, data distribution and a wide range of overlapping spatiotemporal data scopes, as well as robust aggregation and analysis of data using different views.

Due to the limited resources of individual WSN nodes it is imperative that each required task optimize resource consumption since wasting a small portion of reserves can greatly affect a node's lifetime. The current required for a radio to be in receive mode is comparable to when transmitting, which is tens of thousand times greater than sleep mode (17 mA in RX; 500 nA in sleep) [5]. As such, network management protocols need to be designed to allow for duty-cycling (placing the radio in sleep mode for a period of time) with minimal complexity. Current state-of-the-art solutions [6, 7, 8] do not, but instead reduce overhead by relying on the inherent nature of wireless media (i.e. all messages sent wirelessly are broadcasted). However, this means that either the radio is always active or a high level of synchronization exists between any node and all of its neighbors. In contrast, NOA does not rely on broadcasting and provides a solution where radios can be duty-cycled with a minimal level of pair-wise synchronization between communicating nodes.

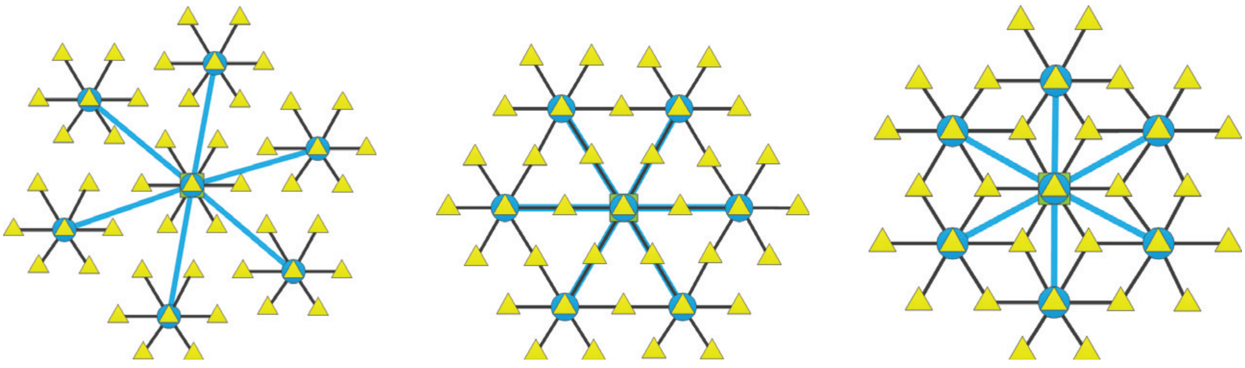
1.2. Prior Work

In-network processing solutions utilize spanning trees, multi-path aggregation, clustering solutions, and hybrid solutions to create a structured network topology [9]. Non-clustering solutions typically follow a paradigm where sink nodes broadcast a message across the network to generate routing paths on which to process data [9, 10]. Typically, data reduction for the non-clustered solutions take an approach where data is stored in situ (i.e. at the node that recorded the data) and the wireless sensor network is treated as a database allowing a data sink to perform queries to search for specific pieces of information (e.g. maximum, minimum, closest value, etc.). Queries are resolved by following the data aggregation backwards, that is each aggregation point forwards the query to the next aggregation point with the closest match to the query. The query is resolved when it reaches the node that recorded the data. This works well for a system where the driver is a centralized node but has its limits when it comes to more in-depth, in-network data analysis as data is not necessarily the same resolution or from neighboring regions. Further, when dealing with a set of data sinks the query resolution mechanism (e.g. spanning tree) needs to be built for each sink. In the case of mobile devices acting as data sinks the process of building the query resolution tree becomes expensive, especially when considering casual connections where a mobile device performs a small number of queries before disconnecting.

Clustering protocols can be categorized in many ways as over-viewed by [6, 11, 12]. Iwanicki categorizes the different clustering protocols into single-level (flat) clustering [13, 14, 15, 16] where nodes are grouped into clusters and each cluster-head reports directly to the sink node. 1-hop clustering hierarchies, [17, 18, 19, 20, 21], use a top-down clustering methodology to reduce the overall communication cost of a network by minimizing each nodes' transmit power. The drawback to this is that all nodes must be able to communicate directly with one another, which limits the physical distance between nodes. These limits lead to distributed bottom-up techniques such as hierarchical beaconing and gossip-based solutions [7, 8, 6]. Du et al. [7] uses hierarchical beaconing where a cluster-head floods a beacon a distance of r-hops (r is the radius of the cluster) to notify children of its existence. Iwanicki and Van Steen [6] use a gossip-based hierarchy maintenance protocol where nodes broadcast a heartbeat (composed of a unique ID, sequence number, hierarchy information, and other merged heartbeats) to their 1-hop neighbors. Nodes merge received heartbeats with their own and broadcast the merged heartbeat during the next period.

2. NOA Approach

Multi-hop, multi-tiered clustering hierarchies provide a scalable network management schema and an in-network data processing paradigm for large scale wireless sensor networks. Autonomous configuration of multi-hop, N-tiered, clustering hierarchies is performed in a bottom-up, distributed fashion. The lowest level nodes, $LeafNodes$ are

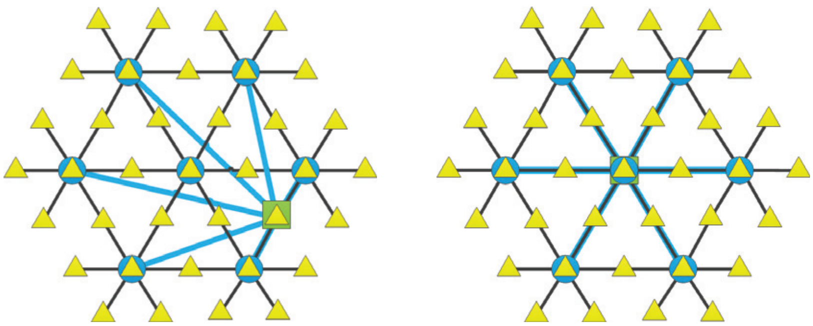


(a) 1-parent (b) 2-parent (c) 3-parent

Figure 1. Yellow triangles represent *LeafNodes*, blue circles are tier 1 cluster-heads, and green squares are tier 2 cluster-heads. Notice how in the 2- and 3-parent clustering hierarchies each node is connected to multiple clusters creating cluster overlap. This overlap allows the proposed solution to share topology information between neighboring clusters and bridges the hierarchical divide allowing a node's data to be analyzed with all of its neighbor's data.

grouped to form a set of tier 1 clusters, tier 1 clusters are grouped to form tier 2 clusters, and so on until the top-most, tier N cluster, requiring only $O(log(n))$ tiers. Current state-of-the-art multi-tiered clustering hierarchies rely on a single parent which divides a network's monitored area into pieces based on the hierarchies' structure (Fig. 1). As such, in-network data processing techniques cannot inherently bridge the gaps and processing decisions are made without all of the local information because the information from two neighboring nodes was sent to different parents for processing. Extending the current single-parent hierarchies to a novel multi-parent clustering hierarchy: bridges the divisions made by single parent clustering, provides a robust solution by removing single points of failure, reduces configuration overhead and reduces the level of synchronization necessary to duty-cycle a node's radio.

Two variants of NOA are proposed. The first, 1-CH, limits the additional responsibilities of a node by only letting a node act as a single cluster-head (i.e. a node can be a parent or a grandparent but cannot be both) and the second, N-CH, removes this constraint allowing a node to act as a cluster-head at each tier (Fig. 2). The proposed clustering algorithm can be partitioned into the following steps: Elevate Node §2.1, Clustering at Tier 1 and Tier n§2.2, and Aggregation of Data §2.3.



(a) 2-parent 1-CH (b) 2-parent N-CH

Figure 2. 1-CH and N-CH 2-parent clustering variants. Yellow triangle nodes are *LeafNodes*. Blue circle nodes are tier 1 cluster-heads. Green squares are tier 2 cluster-heads. Notice in the N-CH version the tier 3 cluster-head is also a tier 2 cluster-head but not in the 1-CH variant. Adding this constraint requires more overhead but reduces the single points of failure and reduces the recovery cost if a high tiered cluster-head is lost.

2.1. Elevate Node

Nodes that are only *LeafNodes* or are *TierNodes* with at least 4 children, and in either case do not have the required number of parents, randomly decide to elevate themselves or an ancestor to the next tier level. After a time period if no node has been elevated the nodes try again until a suitable parent is found. This solution leads to a non-optimal placement of cluster-heads results in more cluster-heads than necessary. Optimization of cluster-head placement can be performed by: making nodes cognitive of their location, using a flat-clustering algorithm such as [16] to create a hexagonal grid overlay, weighting the cluster-head elevation decision such that nodes that are optimally located are selected, or merge and split clusters to minimize cluster-heads after initial configuration. The simulation performed herein assumes that nodes are cognitive of their location. As such each node calculates its cell index in a hexagonal grid overlay and from that decides whether it should elevate to the next tier based on its location in the case of a $LeafNode$ or the cluster center in the case of a $TierNode$ using the following routine for the 2-parent variant.

```
x = horizontal cell index of cluster center;  y = vertical cell index of cluster center
tier = tier level of cluster-head to be elected
mody = pow(2, tier+1);  modx = pow(2, tier)
if((y%mody==1 && x%modx==0) || (y%mody==mody/2+1 && x%modx==modx/2))
  return true;
return false;
```

2.2. Clustering at Tier 1 and Tier n

$LeafNodes$ that are elevated to a tier 1 cluster-head send a new parent notification to their 1-hop neighbors. $LeafNodes$ that receive the parent notification can respond with a parent acceptance that includes a list of parents that the node is currently connected to. It also updates its current parents by sending them information about the newly connected parent. Introducing a delay before sending a response/update gives other parents time to send parent notification messages, thus reducing the number of required parent updates.

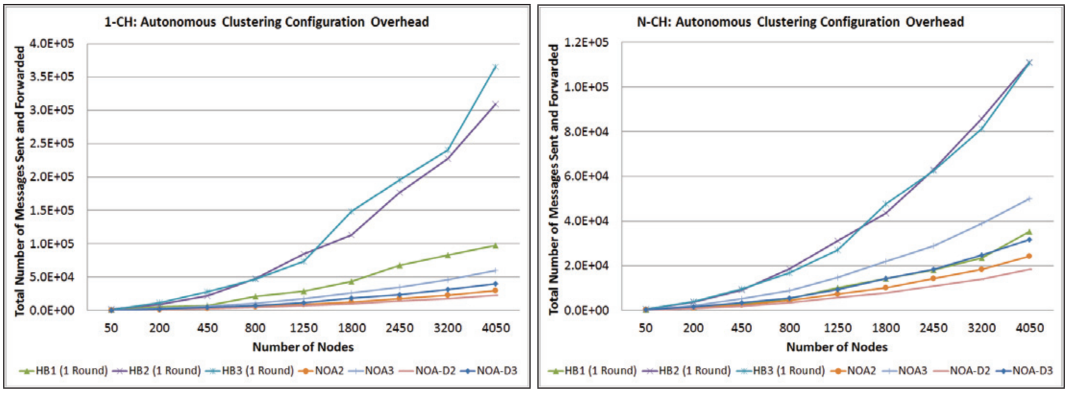
When a $TierNode$ of tier n, tn_n, has at least four children, fewer than the required number of parents, and decides to elevate itself or a descendant as its parent. tn_n if elevating itself becomes a $tn_n + 1$ node or if elevating a descendant starts a random depth-first search of its descendants to find a $LeafNode$ that is not currently acting as a $TierNode$ and elects it as $tn_n + 1$. $tn_n + 1$ then contacts each of its prospective children with a message notifying them that it is a possible parent. The children reply with an acceptance response. As when clustering at tier 1, children are responsible for informing and updating parents of other connected parents.

2.3. Aggregation and Dispersion of Data

As soon as a $LeafNode$ becomes a $TierNode$ the data aggregation and dispersion process starts. $LeafNodes$ send data to their parents, where the data is aggregated and forwarded to the next level parents and so on until the data reaches the highest tiered cluster-heads in the network. The multi-parent solution introduces data redundancy by sending a node's data to all of its parents. Further, because parent nodes are spread across the network, some resolution of data from any node is dispersed network wide. Thus, a data sink that connects to any node in the network only has to search O(log(n)) cluster-heads, where n is the number of nodes in the scope of interest, to find a view of the network covering the area of interest.

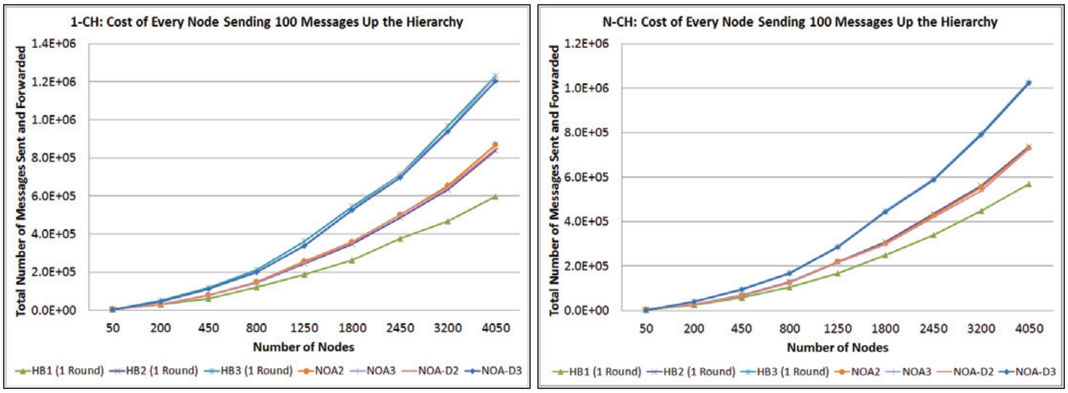
2.4. Simulation

NS-3 was used to simulate 1-CH and N-CH variants of NOA^2 and NOA^3 a 2-, 3-parent hierarchy configuration algorithm, respectively with, $NOA - D$, and without, NOA, a delay imposed before sending parent accept and parent update messages. A single round of 1-, 2-, and 3-parent hierarchical beaconing, HB was also simulated. NS-3's 802.11s mesh networking framework [22] was utilized to simulate a MAC protocol and unreliable wireless radio. The network size was varied from 50 to 4050 nodes with each network size being run 30 times, varying the run number from 1 to 30, and changing the random number generators for each run. A simulation run consisted of configuring the clustering hierarchy and having every node in the network send sensor data up the hierarchy every five seconds. The simulations recorded the overhead in terms of number of messages required to configure the network and the communication costs associated with aggregating data using the different approaches. Note, when simulating hierarchical beaconing a sequence number was attached to each message to ensure a message was forwarded only once by any node otherwise the message explosion created by flooding created communication issues resulting in a fragmented hierarchy.



(a) Nodes limited to acting as only one cluster-head (b) Nodes allowed to be many cluster-heads

Figure 3. Autonomous Clustering Configuration Overhead



(a) Nodes limited to acting as only one cluster-head (b) Nodes allowed to be many cluster-heads

Figure 4. Number of hops required to send 100 messages from each node up the hierarchy.

3. Algorithm Evaluation

The labeling scheme for the different variants is as follows. The proposed solution with and without delay is labeled as $NOA - D$ and NOA respectively. Hierarchical beaconing is HB. Superscripting defines the number of parents. Subscripting defines whether the solution required a node to act as only one cluster-head (1-CH) or as many cluster-heads (n-CH) (Fig. 2).

Fig. 3 compares the hierarchy configuration cost of a single round of hierarchical beaconing, HB to NOA. The NOA variants have significant configuration cost reduction when compared side by side with equivalent HB solutions, and many variants perform better than even the 1-parent HB solution. As such a multi-parent hierarchy can be configured using the NOA algorithm with a lower overhead than using HB to configure a single parent hierarchy while providing the additional benefits of the multi-parent structure. Further, NOA uses MAC level and routing level acknowledgments to provide reliability across an unreliable radio, whereas HB relies on broadcasting and no acknowledgments are used. The 1-CH solutions required more overhead for both configuration and aggregation when compared to the equivalent N-CH solutions because 1-CH has to send all messages up to two times as far to allow nodes on the edge of a cluster to be cluster-heads and the configuration process requires additional messages. While there is an extra cost associated with the 1-CH solution, using it removes another single point of failure thus increasing the robustness of the hierarchical structure.

The data aggregation cost for all of the equivalent 2- and 3-parent solutions was similar, which further validated that the hierarchies configured by each equivalent solution were similar. Interestingly the 2-parent solutions did not require twice as many hops nor did the 3-parent solutions require three times as many. This is because in the multi-parent clustering hierarchies the parents are closer together, allowing the data to be aggregated and reduced earlier. As such, a round robin approach can be taken such that a node only sends data to one parent at a time, rotating the parent with each new value. This would cut the overhead of data aggregation as shown in Fig. 3 by approximately a half or a third depending on the number of parents while still retaining the current benefits provided by multiple parents.

4. Conclusion

Simulation of NOA saw up to a 92% reduction in the configuration overhead required to construct a multi-hop, N-tiered, cluster hierarchy by changing the hierarchy's topology distribution mechanism from flooding, as in hierarchical beaconing, to forwarding through shared children. NOA also reduces the complexity of duty-cycling a node's radio, bridges the hierarchical divisions created when using a single-parent approach, and removes many of the single points of failure in a single-parent hierarchy making the network more robust. Further, unlike current state of the art solutions, NOA uses MAC and routing level acknowledgments for at a minimum the hierarchy configuration messaging. The hierarchy generated using NOA can be used to create scalable, robust, and reliable in-network data analysis solutions while benefiting from multi-path communication, and data redundancy, as well as cluster overlap and varying overlapping spatiotemporal views of the network.

References

[1] M. Jha, T. Sharma, Secure Data aggregation in Wireless Sensor Network: A Survey, International Journal of Engineering Science 3.

[2] P. Andreou, D. Zeinalipour-Yazti, A. Pamboris, P. K. Chrysanthis, G. Samaras, Optimized query routing trees for wireless sensor networks, Inf. Syst. 36 (2) (2011) 267–291. doi:10.1016/j.is.2010.06.001.

[3] M. Xie, S. Han, B. Tian, S. Parvin, Anomaly detection in wireless sensor networks: A survey, Journal of Network and Computer Applications.

[4] S. Bhatti, J. Xu, Survey of target tracking protocols using wireless sensor network, in: Wireless and Mobile Communications, 2009. ICWMC'09. Fifth International Conference on, IEEE, 2009, pp. 110–115.

[5] T. Instruments, CC1101 Low-Power Sub-1GHz RF Transceiver (2010).

[6] K. Iwanicki, M. Van Steen, Multi-hop cluster hierarchy maintenance in wireless sensor networks: A case for gossip-based protocols, Wireless Sensor Networks (2009) 102–117.

[7] S. Du, A. Khan, S. PalChaudhuri, A. Post, A. Saha, P. Druschel, D. Johnson, R. Riedi, Safari: A self-organizing, hierarchical architecture for scalable ad hoc networking, Ad Hoc Networks 6 (4) (2008) 485–507.

[8] S. PalChaudhuri, R. Kumar, R. G. Baraniuk, D. B. Johnson, Design of Adaptive Overlays for Multi-scale Communication in Sensor Networks., in: V. K. Prasanna, S. S. Iyengar, P. G. Spirakis, M. Welsh (Eds.), DCOSS, Vol. 3560 of Lecture Notes in Computer Science, Springer, 2005, pp. 173–190.

[9] E. Fasolo, M. Rossi, J. Widmer, M. Zorzi, In-network aggregation techniques for wireless sensor networks: a survey, Wireless Communications, IEEE 14 (2) (2007) 70–87.

[10] K. il Hwang, J. In, D. S. Eom, Distributed Dynamic Shared Tree for Minimum Energy Data Aggregation of Multiple Mobile Sinks in Wireless Sensor Networks., in: K. Römer, H. Karl, F. Mattern (Eds.), EWSN, Vol. 3868 of Lecture Notes in Computer Science, Springer, 2006, pp. 132–147.

[11] A. A. Abbasi, M. F. Younis, A survey on clustering algorithms for wireless sensor networks., Computer Communications 30 (14-15) (2007) 2826–2841.

[12] O. Boyinbode, H. Le, A. Mbogho, M. Takizawa, R. Poliah, A Survey on Clustering Algorithms for Wireless Sensor Networks., in: T. Enokido, F. Xhafa, L. Barolli, M. Takizawa, M. Uehara, A. Durresi (Eds.), NBiS, IEEE Computer Society, 2010, pp. 358–364.

[13] M. Handy, M. Haase, D. Timmermann, Low energy adaptive clustering hierarchy with deterministic cluster-head selection., in: MWCN, IEEE, 2007, pp. 368–372.

[14] A. Chamam, S. Pierre, A distributed energy-efficient cluster formation protocol for wireless sensor networks, in: Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE, IEEE, 2009, pp. 1–5.

[15] G. S. Tomar, S. Verma, Dynamic Multi-level Hierarchal Clustering Approach for Wireless Sensor Networks., in: D. Al-Dabass (Ed.), UKSim, IEEE, 2009, pp. 563–567.

[16] H. Zhang, A. Arora, GS3: scalable self-configuration and self-healing in wireless sensor networks ., Computer Networks 43 (4) (2005) 459–480.

[17] B. Okeke, K. Law, Multi-level clustering architecture and protocol designs for wireless sensor networks, in: Proceedings of the 4th Annual International Conference on Wireless Internet, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, p. 5.

[18] Y. Jin, L. Wang, Y. Kim, X. Yang, EEMC: An energy-efficient multi-level clustering algorithm for large-scale wireless sensor networks, Computer Networks 52 (3) (2008) 542–562.

[19] B. Nazir, H. Hasbullah, Energy Efficient Multi Hierarchy Clustering Protocol for Wireless Sensor Network (EMHC), ISRN Communications and Networking.

[20] S. Soni, N. Chand, Energy Efficient Multi-Level Clustering To Prolong The Lifetime of Wireless Sensor Networks, CoRR abs/1005.4031.

[21] M. Mamuny, N. Nakaya, Y. Hagihara, G. Chakraborty, et al., HEHC: Heterogeneous-aware enhanced hierarchical clustered scheme for wireless sensor networks, in: SICE Annual Conference (SICE), 2011 Proceedings of, IEEE, 2011, pp. 1517–1522.

[22] K. Andreev, P. Boyko, IEEE 802.11 s Mesh Networking NS-3 Model, in: Workshop on ns3, 2010.